Factsheet on Disclosure Avoidance for the 2010 Demonstration Data Product - Demographic and Housing Characteristics File (v2022-08-25)

The 2010 Demonstration Data Product – Demographic and Housing Characteristics File (DHC) (v2022-08-25) was generated by running 2010 Census data through the 2020 Disclosure Avoidance System's (DAS) TopDown Algorithm. This is the same algorithm used to apply confidentiality protections for the 2020 Census P.L. 94-171 Redistricting Data Summary File that was released in August and September 2021. The Census Bureau has produced these demonstration data to allow data users to evaluate and provide feedback on the potential impact of planned confidentiality protections on the resulting fitness-for-use of the data for their intended uses. Feedback from data users on these demonstration data will inform final parameter settings for production of the 2020 Census DHC data product. Prior iterations of DHC demonstration data were released in October 2019, May 2020, and March 2022.

This release includes the proposed DHC tables. Content includes sex, age, Hispanic origin, race, relationship to householder, group quarters population, household and family type, tenure, and vacancy. Some tables are repeated by major race and Hispanic origin groups. Since the previous demonstration data release, additional tables and geographies were added based on public feedback. Mainly, the lowest level of geography for some tables was changed from state and county level to census tract level and iterated tables were added for sex by single-year of age at the census tract level.

Tuning the DAS for the Demographic and Housing Characteristics File

These demonstration data reflect the Census Bureau's efforts to meet an array of accuracy targets based on external stakeholder feedback and internal program requirements. The Census Bureau worked to meet these targets by making a series of adjustments ("tuning") to the amount of privacy-loss budget (PLB) applied to different sets of tabulations. The Census Bureau established these targets based on subject matter expertise and public feedback. The full breakdown of PLB allocations by individual query and geographic level that reflect this initial tuning is available at: https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/02-Demographic_and_Housing_Characteristics/2022-08-25_Summary_File/2022-08-25_Privacy-Loss_Budget_Allocations.pdf.

Accuracy of the data, in this context, specifically refers to the concept of statistical accuracy. Consistent with the Office of Management and Budget's Statistical Policy Directive #1 (as codified in Title III of the Foundations for Evidence-based Policymaking Act of 2018), statistical accuracy means that the Census Bureau's publicly released data products meet established information quality guidelines while also protecting the confidentiality of respondents' information and, when appropriate, providing information on limitations of the data that may assist data users in determining the suitability of the data for their purposes.

Privacy-Loss Budget

In order to ensure consistency between data products, the DAS uses the full set of tabulations included in the P.L. 94-171 Redistricting Data Summary File as constraints when processing the DHC. As such, the DHC inherits the PLB expended on the redistricting data and supplements this with additional PLB allocated to the tabulations necessary to produce the DHC. A PLB of rho=3.65 (person tables) and rho=6.14 (housing unit tables) was assigned to these new DHC queries. When combined with the PLB allocated to the redistricting data, rho=2.56 (person tables) and rho=0.07 (housing unit tables), this

brings the total cross-product PLB reflected in the DHC demonstration data to rho=12.42. This is equivalent to a global epsilon/delta PLB of epsilon=46.24 with delta=10^{-10}. The full confidentiality protection is reflected in rho, not the conversion to a single point on the epsilon, delta continuum that it summarizes.

It is important to note that the PLB for this demonstration data product is not the final PLB that will be used for the 2020 Census production run of the DHC file; the Census Bureau's Data Stewardship Executive Policy (DSEP) committee will make final PLB decisions for the 2020 Census DHC this fall.

Geographic Hierarchy

To improve the fitness-for-use of the DHC for the Census Bureau's Population Estimates Program and many important demographic use cases of DHC data, the Census Bureau has made changes to the geographic processing hierarchy used within the TopDown Algorithm. Chief among these changes is the incorporation of "Population Estimates Primitive Geographies" into the hierarchy. Population Estimates Primitive Geographies are the most granular geographic unit used by the Census Bureau's Population Estimates Program to derive tables for every geography for which official population estimates are produced. Because the Population Estimates Primitive Geographies do not always align with standard tabulation tracts, the TopDown Algorithm's geographic processing hierarchy also incorporates tract subsets and tract subset groups. Tract subsets are defined as the intersection of the Population Estimates Primitive Geographies with census tabulation tracts. Tract subset groups are the union of multiple tract subsets that are all within the same Population Estimates Primitive Geography. For this demonstration data product, the optimized block group level of the algorithm's geographic hierarchy was further modified to improve accuracy of tabulations for unified school districts. We continue to work on elementary and secondary school districts.

These changes to the TopDown Algorithm's geographic processing hierarchy are used exclusively within the algorithm as part of the disclosure avoidance process. They do not impact how the data are tabulated within the published DHC demonstration data file, which uses traditional tabulation geographical definitions.

How to Submit Feedback

Data user feedback has been instrumental to the iterative development of the DAS. The Census Bureau encourages data users to examine the tables and associated detailed summary metrics to assess whether they are fit for stakeholders' data uses. During this final round of evaluation and feedback, we ask that you identify tables and geographies that have differences that would impact your use of the data and provide acceptable targets when possible.

All comments on these demonstration data should be submitted by September 26, 2022, to 2020DAS@census.gov using the subject "2020 Census Data Products."

Note: The Census Bureau may choose to publish detailed summaries of the comments and feedback received. If so, the Census Bureau will make every effort to remove identifying information, such as names of individuals and organizations.